P1 1111327

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office

January 06, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM
THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK
OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT
APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A
FILING DATE.

APPLICATION NUMBER: *60/420,458*
FILING DATE: *October 23, 2002*
RELATED PCT APPLICATION NUMBER: *PCT/US03/34028*

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

N. WILLIAMS
Certifying Officer

**PRIORITY
DOCUMENT**
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

Approv

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(b)(2).

Docket Number	910000-2033	Type a plus sign (+) inside this box 6	+

INVENTOR(S)/APPLICANT(S)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
Joung	J.	Keith	Molecular Pathology Unit, Massachusetts General Hospital, Boston, MA
Pabo	Carl	O.	Sangamo Biosciences, Inc., Richmond, CA

TITLE OF THE INVENTION (280 characters max.)

CONTEXT SENSITIVE PARALLEL OPTIMIZATION OF ZINC FINGER DNA BINDING DOMAINS

CORRESPONDENCE ADDRESS

Thomas J. Kowalski, Esq.
Frommer Lawrence & Haug LLP
745 Fifth Avenue, 10 Floor
New York, New York 10151/Tel. (212) 588-0800/Fax (212) 588-0500

STATE	New York	ZIP CODE	10151	COUNTRY	USA

ENCLOSED APPLICATION PARTS (Check all that apply)

☒ Specification, Claims and Drawings (Figures 1-11)

48 pages (37 Specification· claims 1-46 on pgs. 31-36 and 1 pg. abstract)

☒ Drawing(s) 11 Sheets

☒ Applicants Qualify for Small Entity Status

☐ Other (specify)

METHOD OF PAYMENT (check one)

☒ A check or money order is enclosed to cover the Provisional filing fees

☒ The Commissioner is hereby authorized to charge filing fees and credit deposit account number. 50-0320

PROVISIONAL FILING FEE AMOUNT(S)	$ 80.00

1

00093549

PROVISIONAL APPLICATION COVER SHEET

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No.
☐ Yes, the name of the U S. government agency and the Government contract number are:

Respectfully submitted,

SIGNATURE	*[signature]*	DATE	October 23, 2002
TYPED OR PRINTED NAME	Thomas J. Kowalski, Esq. Amy Leahy, Ph.D.	REGISTRATION NO. (if appropriate)	32,147 47,739

PROVISIONAL APPLICATION FILING ONLY

EXPRESS MAIL

Mailing Label Number   EV073695357US

Date of Deposit   October 23, 2002

I hereby certify that this paper or fee is being deposited with the United States
Postal Service "Express Mail Post Office to Addressee" Service under 37 CFR
1.10 on the date indicated above and is addressed to Assistant Commissioner
for Patents, Washington, DC 20231

Charles B Jackson

(Typed or printed name of person mailing paper or fee)

[signature]

(Signature of person mailing paper or fee)

2

CERTIFICATE OF MAILING – SEPARATE PAPER

Attorney Docket No. : 910000-2033
Serial No. : TBA
Title : **CONTEXT SENSITIVE PARALLEL OPTIMIZATION OF ZINC FINGER DNA BINDING DOMAINS**
Filed : Herewith
Applicant(s) : Joung, et al.

Date of Deposit: October 23, 2002 via "Express Mail" # EV073695357US

I hereby certify that a Patent Application and related documents for the above-referenced application are being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to:

Assistant Commissioner for Patents
Box Patent Application
Washington, D.C. 20231

Charles B. Jackson
(Typed or printed name of person mailing paper or fee)

(Signature of person mailing paper or fee)

00093646

TITLE OF THE INVENTION

CONTEXT SENSITIVE PARALLEL OPTIMIZATION OF ZINC FINGER DNA
BINDING DOMAINS

RELATED APPLICATIONS/PATENTS & INCORPORATION BY REFERENCE

Each of the applications and patents cited in this text, as well as each document or
reference cited in each of the applications and patents (including during the prosecution
of each issued patent; "application cited documents"), and each of the PCT and foreign
applications or patents corresponding to and/or claiming priority from any of these
applications and patents, and each of the documents cited or referenced in each of the
application cited documents, are hereby expressly incorporated herein by reference.
More generally, documents or references are cited in this text, either in a Reference List
before the claims, or in the text itself; and, each of these documents or references ("herein
cited references"), as well as each document or reference cited in each of the herein-cited
references (including any manufacturer's specifications, instructions, etc.), is hereby
expressly incorporated herein by reference.

STATEMENT OF RIGHTS TO INVENTION MADE UNDER
FEDERALLY SPONSORED RESEARCH

FIELD OF THE INVENTION

The present invention relates to Zinc finger polypeptides having DNA binding
domains, and to methods of selecting Zinc finger polypeptides that bind to sequences of
interest.

BACKGROUND OF THE INVENTION

At any given time, only a fraction of the genes in the genome of an organism are
expressed and/or producing functional protein products. The profile of proteins expressed

1

in an organism varies greatly between cell types and changes over time, depending on factors such as stage of development, stage of the cell cycle and response to environmental factors. Furthermore, gene expression is often mis-regulated in disease.

Gene expression is controlled, in part, by proteins known as transcription factors. The presence of a particular combination of such transcription factors determines whether a gene is switched on or off at any given time and place. Transcription factors are modular proteins. They contain at least one DNA-binding domain (DBD) and one or more effector or regulatory domains. DBDs act as targeting devices to localize transcription factors to specific sequences or "target sites" on the chromosomal DNA. Effector domains function to direct the localization of specific activities to a gene or locus of interest, ultimately enabling transcription of that gene to be up- or down regulated.

The ability to artificially manipulate gene expression has enormous potential for biological research and for the development of new agents for gene therapy. Realizing this potential requires the ability to engineer DNA binding domains that recognize "target site" sequences with high affinity and specificity. Many DNA-binding proteins contain independently folded domains for the recognition of DNA, and these domains in turn belong to a large number of structural families, such as the leucine zipper, the "helix-turn-helix" and zinc finger (Zf) families. Most sequence-specific DNA-binding proteins bind to the DNA double helix by inserting an α-helix into the major groove (Pabo and Sauer 1992 Annu. Rev. Biochem. 61:1053-1095; Harrison 1991 Nature (London) 353: 715-719; and Klug 1993 Gene 135:83-92). Sequence specificity results from the geometrical and chemical complementarity between the amino acid side chains of the a-helix and the accessible groups exposed on the edges of base-pairs. In addition to this direct reading of the DNA sequence interactions with the DNA backbone stabilize the complex and are sensitive to the conformation of the nucleic acid, which in turn depends on the base sequence (Dickerson and Drew 1981 J. Mol. Biol. 149:761-786)

Zfs have become the DBD of choice in efforts to engineer custom-made transcription factors. A Zf is an independently folded zinc-containing mini-domain, the structure of which is well known in the art and defined in, for example, Miller et al., (1985) EMBO J. 4:1609; Berg (1988) Proceedings of the National Academy of Sciences

2

(USA) 85:99; Lee et al., (1989) Science 245:635 and Klug, (1993) Gene 135:83. The crystal structures of Zf DNA complexes show a semi-conserved pattern of interactions, in which typically 3 amino acids from the a-helix of the Zf contact 3 adjacent base pairs (bp) or a "subsite" in the DNA (Pavletich et al., (1991) Science 252:809; Fairall et al., (1993) Nature 366:483; and Pavletich et al., (1993) Science 261:1701). Thus, the crystal structure of Zif268 suggested that Zf DBDs might function in a modular manner with a one-to-one interaction between a Zf and a 3 bp "subsite" in the DNA sequence. In naturally occurring transcription factors, multiple Zfs are typically linked together in a tandem array to achieve sequence-specific recognition of a contiguous DNA sequence (Klug, (1993) Gene 135:83).

Multiple studies have shown that it is possible to artificially engineer the DNA binding characteristics of individual Zfs by randomizing the amino acids at the α-helical positions involved in DNA binding and using selection methodologies such as phage display to identify desired variants capable of binding to DNA target sites of interest (Rebar et al., (1994) Science 263:671; Choo et al., (1994) Proceedings of the National Academy of Sciences (USA) 91:11163; Jamieson et al., (1994) Biochemistry 33:5689; Wu et al., (1995) Proceedings of the National Academy of Sciences (USA) 92: 344). Furthermore, by fusing such recombinant Zf DBDs to regulatory or effector domains, it has been possible to artificially regulate expression of transfected reporter genes in cultured cells. For example, Beerli et al., (Beerli et al., (1998) Proceedings of the National Academy of Sciences (USA) 95:14628) reported construction of a chimeric·six finger Zf protein fused to either a KRAB, ERD, or SID transcriptional repressor domain, or the VP16 or VP64 transcriptional activation domain. This chimeric Zf protein was designed to recognize an 18 bp target site in the 5' untranslated region of the human erbB-2 gene. Using this construct, the authors were able to either activate or repress a transiently expressed reporter luciferase construct linked to the erbB-2 promoter.

Further studies have demonstrated that such recombinant Zf transcription factors can also be used to regulate expression of endogenous genes in their native chromosomal context (Reik et al., (2002) Current Opinions in Genetics & Development 12:233). Clinically relevant human genes that have been successfully regulated in this way include MDR1, erythropoietin, erbB-2 and erbB-3, VEGF, and PPARgamma. In the case of

3

VEGF (Liu et al., (2001) Journal of Biological Chemistry 276:11323), proportional up-regulation by the designed transcription factor of all three distinct splice isoforms generated by this locus was observed, illuminating the utility of endogenous gene control in therapeutic settings (proper isoform ratio is essential for the proangiogenic function of VEGF). In the case of PPARgamma, use of a transcriptional repressor designed to downregulate the expression of two PPARgamma isoforms allowed "mutation-free reverse genetics" analysis that illuminated a unique role for the PPARgamma2 isoform in adipogenesis (Ren et al., (2002) Genes &Development 16:27).

The vast majority of methods used to produce custom–designed Zf DBDs utilize large Zf libraries in which the key amino acids required for DNA binding have been randomized. To select Zfs with the desired DNA binding characteristics from such libraries most researchers use phage display technology, in which the proteins encoded by the Zf library are expressed on the surface of the bacteriophage. Phage particles displaying Zf motifs with the desired sequence specificity are identified using standard techniques that select on the basis of DNA binding affinity and specificity and are then subjected to multiple rounds of selection and amplification. Rebar and Pabo (Rebar et al., (1994) Science 263:671) first used this method to produce a recombinant version of Zif268 with altered DNA-binding specificity.

More recently a bacterial "two-hybrid" method has been developed. In this system Zf-DNA interactions are required for cell growth and survival (Joung et al., (2000) Proceedings of the National Academy of Sciences (USA) 97:7382 and US Patent Application No. 20020119498). The bacterial two-hybrid system has an extremely low background rate and, because it does not require multiple rounds of selection and amplification, it is significantly faster to perform than phage display methods. Furthermore, the bacterial two-hybrid system has an added advantage in that, unlike phage display, the Zf-DNA binding interaction occurs within living cells. Thus, Zfs identified using this method are more likely to function reliably in a cellular context. Joung et al. (Joung et al., (2000) Proceedings of the National Academy of Sciences (USA) 97:7382) demonstrated that Zf candidates selected using this method were at least as effective as those selected for binding to the same DNA targets using phage display.

In order to use recombinant Zfs to target a gene of interest within the genome, the

target site sequence recognized should be sufficiently long that statistically it occurs only once in the genome. In the case of the human genome, a multi-finger Zf protein recognizing a stretch of about 16 bp or more should be generated for this to be achieved (Liu et al., (1997) Proceedings of the National Academy of Sciences (USA) 94:5525). Statistically, assuming random base distribution, a unique 16 bp sequence will occur only once in $4.3x\ 10^9$ bp, thus a 16 bp sequence should be sufficient to specify a unique address within the approximately 3.5×10^9 bp that make up the human genome (Liu et al., (1997) Proceedings of the National Academy of Sciences (USA) 94:5525). Similarly, an 18 bp address specified by a six finger protein, would enable sequence specific targeting within 6.8×10^{10} bp of DNA. Such a six-finger protein would thus be able to uniquely specify any locus within all currently known genomes.

At present there are three main methods by which such multi-finger Zf proteins can be selected from a library and produced. These are known as the parallel selection, sequential selection and bipartite selection methods (for review, see Beerli and Barbas, (2002) Nature Biotechnology 20:135).

The basic assumption of parallel selection is that individual Zf domains are functionally independent and can therefore be recombined with one another to recognize any desired DNA sequence. Thus, individual fingers selected to bind to any given 3 bp subsite can be "stitched" together to produce a multi-finger DBD. Although several multi-finger proteins have been produced using this method (including Desjarlais et al., (1993) Proceedings of the National Academy of Sciences (USA) 90:2256; Choo et al., (1994) Nature 372:642), a major limitation arises from the oversimplified model on which it is based, i.e., that Zfs bind DNA as independent modular units. In reality, differences in the amino acid sequence of one Zf can affect the function of neighboring fingers. In other words, there exists in some natural Zf proteins the propensity for necessary interaction between individual Zf domains, or "positions," termed finger "context dependence" or "position sensitivity." For example, inter-finger contacts have been reported in the crystal structures of synthetic zinc finger proteins selected to bind to a TATA box sequence (Wolfe et al., (2001) Structure 9:717).

In addition, it has been noted that some Zfs display "target-site overlap," in which zinc finger domains work cooperatively to recognize DNA sequence at their subsite

junctions (Pavletich et al., (1991) Science 252:809; Elrod-Erickson et al., (1996) Structure 4:1171; Kim et al., (1996) Nature Structural Biology 3:940; Isalan et al., (1997) Proceedings of the National Academy of Sciences (USA) 94:5617). Thus, although the parallel screening method can identify functional multi-finger DBDs, ignoring the importance of finger context may produce sub-optimal multi-finger proteins.

The sequential selection method was developed by Greisman and Pabo (Greisman et al., (1997) Science 275:657 and US Patent No. 6,410,248) in an attempt to address the lack of context dependence that plagues the parallel selection method. In this method, DNA-binding specificities of individual Zf domains are altered sequentially in the context of the other Zfs. Thus, finger three of a three-finger protein is replaced by a finger one in which the critical amino acid residues have been randomized. This library is then selected in the context of the two original fingers, which serve as anchors. After selection, the N-terminal anchor finger is removed and a finger two library is attached to the C-terminus. Selection of this library ensures that the new finger two works well in the context of the finger one selected in the previous round. In the final step, the last remaining anchor finger is discarded and a randomized finger three is attached to the C-terminus, again followed by selection. In this manner, each finger of the new three-finger protein is selected in the context of its neighboring finger, preventing problems associated with target site overlap. Recently the crystal structure of a sequentially selected protein in complex with its TATA box target sequence has been reported (Wolfe et al., (2001) Structure 9:717). Although sequential selection undoubtedly overcomes the problems associated with the parallel selection method, the need to sequentially generate multiple Zf libraries for each protein produced makes this a very labor- and time-intensive procedure and therefore, not suitable for repeated or high-throughput use.

The most recently developed Zf selection protocol is the bipartite method. This technique was developed by Isalan et al. (Isalan et al., (2001) Nature Biotechnology19: 656) with the aim of combining the advantages of the parallel and sequential methods but avoiding the context sensitivity problems of the parallel selection method. Bipartite selection makes use of a pair of prefabricated libraries, each having one-and-a-half fingers of the three Zf protein Zif268 randomized. Selection of these two libraries is carried out in parallel against DNA sequences in which either the first or the last 5 bp of

6

the 9 bp Zif268 target site are exchanged against a target site of interest. After phage display selection, pools of binding fingers from the two prefabricated libraries are recombined to produce a partially selected library of three finger proteins. Further rounds of selection are then performed against the full 9 bp sequence of interest. Isalan et al. (Isalan et al., (2001) Nature Biotechnology19:656) used this method to select three finger proteins that bind to sequences within the HIV-1 promoter and found that the proteins produced had affinities comparable to those of Zfs produced using the parallel and sequential strategies.

Thus, the bipartite method avoids target site overlap and position sensitivity problems associated with parallel selection, and also avoids the multiple library production problem associated with sequential selection. However, these benefits have been achieved at the expense of combinatorial diversity. The need to randomize 8 to 10 amino acids within each one-and-a-half finger library presents a combinatorial problem beyond the capability of existing library methods, if significant randomization of the residues is permitted. In an attempt to overcome this defect, Isalan et al. designed the two libraries used in the initial selection to limit the number of amino acid variations. However, this "pre-selection" at the level of the starting libraries means that the full range of possible Zfs are not screened and thus optimal fingers may not even be present in the original libraries.

Although several techniques exist for selecting multi-finger proteins, each of these methods has limitations. An ideal multi-Zf selection strategy would involve one or more, or preferably all of the following elements:

a) retaining maximal combinatorial diversity in the Zf libraries used,

b) avoiding prior assumptions about the role of particular amino acids in binding,

c) overcoming the problems of target-site overlap and position sensitivity,

d) screening and selecting of full length assembled multi-finger Zf proteins directly against the sequence of interest,

e) avoiding post-selection assembly of individual Zfs or groups of Zfs,

f) allowing selection of Zfs which bind to their target sites in a cellular context, and

g) simplifying and expediting procedures for use in high-throughput applications.

Prior to the development of the methods described herein, no strategy was known

7

to combine all of these features.

OBJECT AND SUMMARY OF THE INVENTION

The present invention provides methods for rapidly selecting multi-finger Zf polypeptides that bind to any desired sequence of interest comprising a target site, termed "context sensitive parallel optimization" (CSPO). CSPO overcomes the problems of target site overlap and context sensitivity associated with other methods, without sacrificing combinatorial diversity. A schematic illustration of a CSPO strategy is provided in Figure 1. CSPO uses master libraries in which up to 20 amino acids can be represented at each of the sites randomized within a single Zf, and requires the construction of only one new "secondary" library for each multi-finger polypeptide constructed. In addition, CSPO allows for efficient screening and selection of pre-assembled multi-finger Zf polypeptides having the desired DNA sequence specificity. Methods of the present invention can be used in conjunction with the classical systems known in the art for Zf selection, such as phage-display or polysome systems. Preferably, methods of the present invention can be used in conjunction with prokaryotic or eukaryotic cell-based selection methods (e.g. a bacterial, yeast or mammalian two-hybrid systems), thus ensuring that a multi-finger polypeptide selected functions well in a cellular context. In summary, the methods of the present invention provide a rapid and feasible means to select optimized multi-finger proteins with high affinity and specificity.

Accordingly, in one aspect, the present invention provides a method of selecting a Zf polypeptide that binds to a sequence of interest comprising a target site having at least one subsite, wherein the method comprises the steps of:

a) first obtaining primary libraries comprising polypeptides having one variable finger and at least one anchor finger, wherein said variable finger corresponds to a zinc finger of said multi-finger zinc finger polypeptide;

b) incubating said primary libraries with said target site under conditions sufficient to form binding complexes;

d) isolating pools comprising nucleic acid sequences encoding polypeptides, wherein said polypeptides comprise said binding complexes;

e) recombining said pools to produce a secondary library;

f) incubating said secondary library with the sequence of interest under

8

conditions sufficient to form a high-affinity binding complex; and

g) isolating nucleic acid sequences encoding multi-finger zinc finger polypeptides, wherein said polypeptides comprise said high-affinity binding complexes.

The composition of the primary libraries, which are carefully controlled to maintain combinatorial diversity, coupled with the composition of the secondary libraries, which are carefully controlled to account for finger position sensitivity, results in the improved selection of Zf proteins.

These and other objects and embodiments are described in or are obvious from and within the scope of the invention, from the following Detailed Description.

BRIEF DESCRIPTION OF THE DRAWINGS

In the following Detailed Description and Examples reference will be made to the accompanying drawings, incorporated herein by reference.

Figure 1 provides a schematic representation of the required components and steps of the context-sensitive parallel optimization (CSPO) Zf selection strategy that is the object of the present invention.

Figure 2 provides a schematic representation of the PCR-mediated recombination protocol used to generate the secondary libraries used in CSPO.

Figure 3 shows the characterization of a CSPO-selected finger by EMSA and the measurement of the K_D for binding to its specific target.

Figure 4 shows the characterization of a CSPO-selected finger by EMSA and the measurement of the K_D for binding to non-specific DNA.

Figure 5 provides a schematic representation of multi-finger proteins, previously selected by other methods, that were compared to the multi-finger proteins selected using methods of the present invention.

Figure 6 depicts sequences of BCR-ABL target-binding Zfs selected using methods of the present invention, and their activity in bacterial reporter gene expression assays.

Figure 7 depicts binding affinities and specificities (determined using EMSAs) for BCR-ABL target-binding Zfs.

Figure 8 depicts sequences of erb-B2 target-binding Zfs selected using methods of the present invention, and their activity in bacterial reporter gene expression assays.

Figure 9 depicts binding affinities and specificities (determined using EMSAs) for erb-B2 target-binding Zfs.

Figure 10 depicts sequences of HIV-1 promoter-binding Zfs selected using methods of the present invention, and their activity in bacterial reporter gene expression assays.

Figure 11 depicts binding affinities and specificities (determined using EMSAs) for HIV-1 promoter-binding Zfs.

DETAILED DESCRIPTION OF THE INVENTION

I. Introduction

The present invention provides methods for the selection of multi-finger Zf polypeptides that bind to a sequence of interest comprising a target site. Preferably, all of the constituent fingers of the Zf polypeptide are maximally randomized and selected simultaneously for binding to a given sequence of interest. Such a Zf selection strategy advantageously avoids position sensitivity problems while retaining the greatest possible diversity of fingers from which to perform efficient selection.

Other methods known in the art either reduce library variability to within manageable limits, thereby sacrificing combinatorial diversity (e.g. the bipartite selection strategy described above), or require "stitching" together of individually selected Zfs, thereby sacrificing context-sensitivity (e.g. the parallel selection strategy described above). To date, the only selection strategy developed that does not sacrifice combinatorial diversity or position sensitivity, is the sequential selection method described by Greisman and Pabo (Greisman and Pabo (1997) Science 275:657 and US Patent No. 6,410,248). However, the generation of a three finger protein by Greisman and Pabo's sequential selection requires the generation and selection of at least two and preferably three Zf libraries for each protein produced (Wolfe et al., (1999) Journal of Molecular Biology 285: 1917). Because these libraries depend upon the results of a previous selection step, each of these libraries must be produced sequentially. As a result, Greisman and Pabo's sequential selection is comparatively labor- and time-intensive, and

10

therefore, less suitable for routine or high-throughput use.

The present invention provides a Zf selection method that allows maximal combinatorial diversity to be maintained and also allows efficient selection of assembled multi-finger polypeptides directly against their given target site. The method, referred to as context-sensitive parallel optimization or CSPO, achieves this goal by combining two selection/screening steps. The initial selection utilizes primary Zf libraries in which maximal library diversity is maintained. In the second selection/screening step, full length assembled multi-finger Zf proteins are screened directly against the sequence of interest to identify those multi-finger polypeptides that work in a coordinated fashion to give optimal target site binding. This second step essentially selects for fingers that work well together, thereby accounting for finger position sensitivity. No additional post-selection assembly of individual Zfs (or groups of Zfs) is required. Thus, methods of the present invention avoid problems of position sensitivity and target site overlap suffered by other methods known in the art. Furthermore, only one custom-made primary library is needed for each new Zf polypeptide to be selected, thus making methods of the present invention simpler and faster to perform than, for example, the sequential selection method.

The library and selection methods described herein can be used in conjunction with suitable expression and selection methods known in the art. Preferably bacterial two-hybrid selection or some other prokaryotic or eukaryotic cell-based selection method is used. Use of such cell-based methods has the advantage of selecting for Zf-DNA interactions in living cells and therefore, selecting for polypeptides that will function well in a cellular context. In addition, cell-based selection methods are highly efficient to perform. Methods of the present invention can be used with other commonly used Zf expression/selection systems, such as phage display or polysome display, if desired.

II. Definitions

As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

In this disclosure, "comprises," "comprising," "containing" and "having" and the like can have the meaning ascribed to them in U.S. Patent law and can mean " includes," "including," and the like; "consisting essentially of" or "consists essentially" likewise has

the meaning ascribed in U.S. Patent law and the term is open-ended, allowing for the presence of more than that which is recited so long as basic or novel characteristics of that which is recited is not changed by the presence of more than that which is recited, but excludes prior art embodiments.

The term "zinc finger" or "Zf" refers to a polypeptide having DNA binding domains that are stabilized by zinc. The individual DNA binding domains are typically referred to as "fingers." A Zf protein has at least one finger, preferably two fingers, three fingers, or six fingers. A Zf protein having two or more Zfs is referred to as a "multi-finger" or "multi-Zf" protein. Each finger typically comprises an approximately 30 amino acid, zinc-chelating, DNA-binding domain. An exemplary motif characterizing one class of these proteins is -Cys-(X) (2-4)-Cys-(X) (12)-His-(X) (3-5)-His (SEQ ID NO:1), where X is any amino acid, which is known as the "C(2)H(2)class." Studies have demonstrated that a single Zf of this class consists of an alpha helix containing the two invariant histidine residues co-ordinated with zinc along with the two cysteine residues of a single beta turn (see, e.g., Berg and Shi, Science 271:1081-1085 (1996)).

A Zf protein binds to a nucleic acid sequence of interest comprising a "target site." A "target site" is a nucleic acid sequence recognized by a Zf protein. Each finger binds from about two to about five base pairs within the target site, preferably three or four base pairs (the "subsite"). Accordingly, a "subsite" is a subsequence of the target site, and corresponds to a portion of the target site recognized by a single finger. A single Zf preferably recognizes a 3 or 4 bp subsite. A "multi-subsite" is a subsequence of the target site comprising at least 4 bp, preferably 6 bp or more. The target site for a multi-Zf protein comprises at least two, typically three, four, five, six or more subsites or multi-subsites, (i.e., one for each finger of the protein).

" K_D" refers to the dissociation constant for binding of one molecule to another molecule, i.e., the concentration of a molecule (such as a Zf protein), that gives half maximal binding to its binding partner (such as a DNA target sequence) under a given set of conditions. The K_D provides a measure of the strength of the interaction between two molecules, or the "affinity" of the interaction between two molecules. Two molecules that bind strongly to each other have a "high affinity" for each other, while molecules that bind weakly to each other have a "low affinity" for each other.

12

The term "recombinant" when used herein with reference to portions of a nucleic acid or protein, indicates that the nucleic acid comprises two or more sub-sequences that are not found in the same relationship to each other in nature. For instance, a nucleic acid that is recombinantly produced typically has two or more sequences from distinct genes or non-adjacent regions of the same gene, synthetically arranged to make a new nucleic acid sequence encoding a new protein, for example, a DBD from one source and a regulatory or effector region from another source, or a Zf from the native Zif268 protein and a Zf selected from a library. The term "recombination" as used herein, refers to the process of producing a recombinant protein or nucleic acid by standard techniques known to those skilled in the art, and described in, for example, as Sambrook et al., Molecular Cloning; A Laboratory Manual 2d ed. (1989).

"Nucleotide" refers to a base-sugarphosphate compound. Nucleotides are the monomeric subunits of both types of nucleic acid molecules, RNA and DNA. Nucleotide refers to ribonucleoside triphophates, rATP, rGTP, rUTP and rCTP, and deoxyribonucleoside triphosphates, such as dATP, dGTP, dTTP, and dCTP.

"Base" refers to the nitrogen-containing base of a nucleotide, for example adenine (A), cytidine (C), guanine (G), thymine (T), and uracil (U). "Base pair" or "bp" refers to the partnership of bases within the DNA double helix, whereby typically an A on one strand of the double helix is paired with a T on the other strand and a C on one strand of the double helix is paired with a G on the other strand.

"Nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O- methyl ribonucleotides, peptide-nucleic acids (PNAs). Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. The term nucleic

13

acid is used interchangeably with gene, cDNA and nucleotide. The nucleotide sequences are displayed herein in the conventional 5' to 3' orientation.

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an analog or mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. Polypeptides can be modified, e.g., by the addition of carbohydrate residues to form glycoproteins. The terms "polypeptide," "peptide" and "protein" include glycoproteins, as well as non-glycoproteins. The polypeptide sequences are displayed herein in the conventional N-terminal to C-terminal orientation.

The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., a carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine, and methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid. The terms "amino acid residue" or "residue" refer to a specific amino acid position within a polypeptide or protein.

Degenerate codon substitutions or "doping strategies" may be achieved by generating sequences in which any position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer et al., Nucleic Acid Res. 19:5081 (1991); Ohtsuka et al., J. Biol. Chem. 260:2605-2608 (1985); Rossolini et al., Mol. Cell. Probes 8:91-98 (1994)). Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine.

14

Thus, at every position where an alanine is specified by a codon in an amino acid herein, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

The term "library" as used herein refers to a population of nucleic acid sequences that encode Zf polypeptides. Such "libraries" are used in the present invention to screen for and identify Zf polypeptides having desired characteristics from large and complex pool of Zf polypeptides. Such libraries can be created in cell free systems or within eukaryotic cells, prokaryotic cells or viral particles. The term "primary library" refers to a library that has not been enriched for nucleic acids encoding Zf polypeptides with particular characteristics. The term "secondary library" refers to a library that is enriched for nucleic acids encoding Zf polypeptides with particular characteristics.

The term "randomized" or "randomize" refers to a pool of Zf molecules, or the generation of a pool of Zf molecules, in which one of a multitude of possible amino acids is represented at one or more given "variable" amino acid positions. The term "maximally randomized" as used herein, means that the maximum number of different amino acids are represented at the variable amino acid positions. The maximum number of amino acids that can be represented in any given randomized protein is a function of both the number the of variable positions and the maximal diversity of the library system used. Preferably, the maximum number of different amino acids represented at a given variable amino acid position is 20, 16 or most preferably, 19.

"Specific" or "specific-binding" as used herein, refers to the interaction between a protein and a nucleic acid wherein the protein recognizes and interacts with a defined nucleotide sequence, as opposed to a "non-specific" interaction wherein the protein does not require a defined nucleotide sequence to associate with the nucleic acid molecule (for

15

example, a protein that interacts with the phosphate-sugar backbone of the DNA but not the bases of the nucleotides). The strength of the association between the protein and the nucleic acid molecule can vary significantly between different "binding complexes." A "binding complex," as used herein, comprises an association between a sequence of interest, target site or subsite and a Zf binding domain. "Binding complexes" can comprise both weakly-bound Zf proteins and nucleic acids and strongly-bound Zf proteins and nucleic acids. The strength or "affinity"of the association of a Zf with an intended or specified sequence of interest, target site or subsite is expressed in terms of the K_D., as defined above.

"Conditions sufficient to form binding complexes" refers to the physical parameters selected for a binding reaction or "incubation" between a nucleic acid and a protein sample that potentially contains an unknown nucleic acid-binding protein, such as, buffer ionic strength, buffer pH, temperature, incubation time, and the concentrations of nucleic acid and protein, where such physical parameters allow nucleic acids to bind to proteins. Such conditions can be "low-stringency conditions", which are conducive to the formation of "binding complexes" comprising both weakly- and strongly-bound proteins and nucleic acids or "high-stringency conditions", which are conducive to the formation of "high affinity binding complexes" comprising only strongly-bound proteins and nucleic acids. Low-stringency conditions typically comprise high salt concentration and a temperature ranging between 37C and 47C. When DNA-protein "binding reactions" or "incubations" are performed *in vitro*, high-stringency conditions typically comprise lower salt concentrations, a temperature of 65C or greater, and a detergent, such as sodium dodecylsulfate (SDS) at a concentration ranging from about 0.1% to about 2%. When DNA-protein "binding reactions" or "incubations" are performed within living cells, the stringency of the binding reaction is controlled as described by Joung et al. (Joung et al., 2000, Proceedings of the National Academy of Sciences (USA) 97:7382 and US Patent Application No. 20020119498).

Further definitions are provided in context below.

III. Construction of Primary Libraries

The CSPO strategy employs construction and/or use of a separate primary library for each Zf position of the multi-finger protein to be generated. For example, if a two-

finger protein is required, two primary libraries are be produced, the first library having Zf position 1 (the N-terminal Zf) randomized and Zf position 2 held constant as an "anchor" finger. The second primary library would have Zf position 2 (the C-terminal Zf) randomized and Zf position 1 held constant as an "anchor." Primary Zf libraries with 2, 3, 4, 5, 6 or more Zfs can be produced according to the same scheme, with only one Zf position randomized in each library and the remaining fingers held constant to act as "anchors." These primary libraries account for position sensitivity, and are termed "position sensitive," because Zfs are selected using the primary library in which the randomized Zf occurs in the same position relative to the other Zfs, as is required in the final multi-Zf product.

In the Examples given below, three-finger Zf proteins were selected and thus three separate position sensitive primary libraries were used. In "primary library 1" the N-terminal Zf (Zf 1) was randomized while Zf 2 and Zf 3 were held constant. Accordingly, Zf 1 in primary library 1 is the "variable finger" while Zf 2 and Zf 3 each serve as an "anchor finger" and, randomized Zf 1 in primary library 1 is said to "correspond" to the "finger position" of original Zf 1. In "primary library 2" the middle Zf (Zf 2) was randomized while Zf 1 and Zf 3 were held constant. In "primary library 3" the C-terminal Zf (Zf 3) was randomized while Zf 1 and Zf 2 were held constant.

Primary libraries, thus described, do not have to be generated anew for each Zf protein to be selected. "Master" primary libraries can be obtained for selection of any Zf protein having the same number of Zfs. For example, any three-finger Zf protein can be selected using the three-finger "master" libraries outlined above.

The constant "anchor" fingers (and the variable fingers to be randomized as described herein) for the primary library can be taken from any natural or synthetic Zf protein known in the art. The only requirement is that a target site for each of the anchor fingers is available (described below). Typically, constant Zfs are made from any suitable C(2)H(2) Zf protein, such as SP-1, SP-1C, TFIIIA, GLI, Tramtrack, YY1, or ZIF268 (see, e.g., Jacobs, EMBO J. 11:4507 (1992); Desjarlais and Berg, Proc. Natl. Acad. Sci. U.S.A. 90:2256-2260 (1993)). More preferably, the "anchor" Zfs are taken from the naturally occurring Zif268 protein, which are well known in the art and bind strongly to their native target sites. More preferably still, for the given invention, the anchor fingers

17

are the previously phage-selected fingers described by Choo et al. (1994, Nature 372: 642). These fingers were synthetically derived from the Zif268 fingers and are not naturally occurring Zfs. The recognition helices (positions −1, +1, +2, +3, +4, +5,, and +6) of these phage-selected fingers have the sequences DRSSLTR (SEQ ID NO:2) for finger 1, QGGNLVR (SEQ ID NO:3) for finger 2, and QAATLQR (SEQ ID NO:4) for finger 3, and bind to the DNA subsites GCC (SEQ ID NO:5) for finger 1, GAA (SEQ ID NO:6) for finger 2, and GCA (SEQ ID NO:7) for finger 3, respectively. Preferably, the above phage-selected fingers are used in methods of the present invention because they have lower affinity for their subsites than the naturally occurring Zif268 fingers. Without being bound by theory, it is believed that by using low affinity binding Zfs as anchors, it is possible to enforce greater affinity and specificity on the finger being randomized and selected. When multi-finger proteins are selected using strong "anchor" fingers (for example, Joung et al., (2000) Proceedings of the National Academy of Sciences (USA) 97:7382), the recognition helix sequences of proteins typically selected, yield helices that would be predicted to recognize only two out of the three bases in the target subsite. In contrast, by using weaker or lower affinity "anchor" fingers, it is possible to enforce selection of fingers that would be predicted to recognize all three bases in the subsite.

The "variable" finger in each primary library can be based on any naturally occurring or synthetic Zf protein, as for the "anchor" fingers. A "variable" finger comprises randomized amino acids at one or more residue positions of the α-helix. A "variable" finger, as used herein, does not comprise partial or fragmented finger configurations, such as a one-and-a-half finger configuration. Preferably, six amino acid residues in the α-helix of the Zf are randomized. More preferably still, the six amino acid residues at positions −1, +1, +2, +3, +5 and +6 in the α-helix are randomized. Preferably, the variable finger is based upon the Zfs from Zif268. Both variable fingers and anchor fingers can bind to subsites within the target site.

The number of randomized amino acids at a single residue position can be varied up to the maximum limits of the library expression and selection system used. Preferably, all 20 naturally occurring amino acids are represented in any given randomized residue position. Perhaps more frequently, it will be desirable to limit the number of variable amino acids in any given residue position to 19. If cysteine is excluded, the remaining 19

naturally occurring amino acids can be encoded by 24 codons as a result of codon doping schemes wherein some of the codons used encode several amino acids (Wolfe et al., (2001) Structure 9:717). Libraries with 24 codon variations at six variable positions of an α-helix have a diversity of 24^6. A library of such a size is within the limits of known expression and selection systems, such as the bacterial two-hybrid system and phage display. Thus, in one embodiment, methods of the present invention comprise the use of libraries in which 19 different naturally occurring amino acids are represented at one or more variable residue positions of the α-helix. In this instance, the naturally occurring amino acid cysteine is excluded because cysteine can not readily be incorporated into a 24-codon doping strategy.

In yet another embodiment, 16 naturally occurring amino acids are represented in any given randomized residue position within the α-helix. 16 amino acids can also be encoded by 24 codons using codon-doping strategies (see Joung et al., (2000) Proceedings of the National Academy of Sciences (USA) 97:7382). Thus, as for the 19 amino acid library described above, such a 16 amino acid Zf library also has a diversity of 24^6. In the embodiment where a 16 amino acid/24 codon library is used, the excluded amino acids are preferably phenylalanine, tryptophan, tyrosine, and cysteine.

The primary libraries described herein can be synthesized using any known randomization strategy (see for example Joung et al., (2000) Proceedings of the National Academy of Sciences (USA) 97:7382). Such strategies are well known to those skilled in the art and include, for example, the use of degenerate oligonucleotides, use of mutagenic cassettes and techniques based on error prone PCR. Standard recombinant DNA and cloning techniques can also be used for library construction and for incorporation of such libraries into appropriate expression and selection systems. Standard recombinant DNA and cloning techniques are well known to those of skill in the art and are described in laboratory text such as, for example, Sambrook et al., Molecular Cloning; A Laboratory Manual 2d ed. (1989), the contents of which are incorporated herein by reference.

IV. Choice of DNA Targets and production of Target sites

In a preferred embodiment, the target site is chosen from a genomic "address" or location that is within or proximal to, for example, a regulated gene ("gene of interest"), such that the sequence is statistically unique enough to occur only once in the genome.

This ability to specify a unique sequence is a function of the length of the target site and the size of the genome or other desired substrate (such as a nucleic acid vector, for example). For example, assuming random base distribution, a unique 16 bp sequence will occur only once in 4.3×10^9 bp, thus a 16 bp sequence should be sufficient to specify a unique address within 4.3×10^9 bp of sequence. Similarly, an 18 bp address would enable sequence specific targeting within 6.8×10^{10} bp of DNA. The unique target site selected can be located anywhere within or proximal to the gene of interest. Wherein the ultimate aim is to generate a synthetic transcription factor to regulate expression of the gene of interest, it is preferable that the chosen target site is within the general vicinity of the promoter and in a region where chromatin architecture will not impede binding of the Zf protein to the target site (see for example, Liu et al., (2001) Journal of Biological Chemistry 276:11323).

Once the desired sequence of interest has been chosen, target sites for use in screening assays can be produced. The CSPO strategy employs construction and/or use of a separate target site for each subsite within the entire target site. For example, if a 6 bp (2 subsite) target site is specified, two target sites are produced. For example, in the first target site subsite 1 (the 5' subsite) would have the sequence of the gene of interest, and subsite 2 (the 3' subsite) would have a defined "anchor" sequence. In the second target site subsite 2 (the 3' subsite) would have the sequence of the gene of interest, and subsite 1 would have a defined "anchor" sequence. DNA target sites with 2, 3, 4, 5, 6 or more subsites can be produced according to the same scheme, with only one subsite having the sequence of the gene of interest and the remaining subsites having the defined "anchor" sequences. These target sites are referred to as "position sensitive" because the subsites having the sequence of the gene of interest are located at the same position relative to the other subsites, as occurs in the true target site within the gene of interest. In a preferred embodiment, these target sites would be positioned upstream of a test promoter for use in the bacterial two-hybrid system (Joung et al., 2000, Proceedings of the National Academy of Sciences (USA) 97:7382 and US Patent Application No. 20020119498).

Such target sites can be synthesized readily using standard molecular biology techniques (for example using restriction digestion of vector DNA, PCR, or automated nucleic acid synthesis). Such techniques are well known to those skilled in the art and are

described in many laboratory texts such as, for example Sambrook et al., Molecular Cloning, A Laboratory Manual 2d ed. (1989).

V. Polypeptide Library Expression and Selection System

As with other Zf selection strategies, CSPO requires an expression system to enable production of the library-encoded Zf proteins, a mechanism for assaying the binding of the library-encoded Zf proteins to the target sites, target subsites and/or sequence of interest, and a means of selecting from the library those Zfs with the desired binding characteristics.

The primary libraries described above can be expressed using any of a variety of protein expression systems known in the art, such as phage display, polysome display, *in vitro* transcription/translation, or expression in eukaryotic or prokaryotic cells. It would be routine for one skilled in the art to incorporate such a library into such an expression system.

Likewise, there are many methods known in the art that would allow the binding of the library-encoded Zf proteins to their DNA target sites and/or sequences of interest, to be measured, such as by phage display, bacterial two-hybrid and ribosome display. Any known protein expression system and any known protein-DNA binding assay could be combined and used to identify library-encoded Zf proteins having the desired binding characteristics.

In a preferred embodiment, a eukaryotic or prokaryotic cell-based expression and selection system is used. Use of such a cell-based system advantageously provides for the selection and expression of proteins inside living cells, thus the Zf proteins identified are likely to function well in a cellular context.

In a more preferred embodiment, a bacterial "two-hybrid" system is used to express and select the Zfs of the present invention. The bacterial two-hybrid selection method has an additional advantage, in that the library protein expression and the DNA binding "assay" occur within the same cells, thus there is no separate DNA binding assay to set up.

The use of bacterial two-hybrid systems to express and select Zf proteins is described in Joung et al., 2000, Proceedings of the National Academy of Sciences (USA) 97:7382 and US Patent Application No. 20020119498, the contents of which are

21

incorporated herein by reference.

Whichever expression and DNA-binding system is used, a key aspect of the present invention is that a separate primary screen and selection is performed for each "Zf/subsite pair" i.e. if the aim is to select a two finger protein that binds to a given 6 bp target sequence, two parallel selections are performed, one for each Zf/subsite pair. For example, in the scheme described above, in primary selection 1, primary library 1 is expressed and screened for binding to DNA target site 1, i.e. primary library 1 and DNA target site 1 comprise a Zf/subsite pair. Similarly, in primary selection 2, primary library 2 is expressed and screened for binding to DNA target site 2. It follows that, if the aim is to select a three finger protein that binds to a given 9 bp target sequence, three parallel selections are performed, one for each Zf/subsite pair. Similarly, if the aim is to select a six finger protein that binds to a given 18 bp target sequence, six parallel selections are performed.

In a preferred embodiment, the stringency of each of the primary selections should be low; such that each selection yields a pool of Zf proteins with target binding affinities that range from low to high. The rationale for this low stringency selection is that there should be no bias towards Zfs that bind tightly to their target subsite at the primary selection stage, because Zfs so identified may not bind tightly to their target subsite in the context of the Zfs selected against the other subsites that make up the full target sequence. Zfs that bind tightly in the context of the "anchor" fingers may not bind tightly in the context of the full target specific Zf protein. Mechanisms for controlling the stringency of DNA binding reactions are known to those of skill in the art and any such mechanism can be used.

VI. Construction of Secondary Partially Optimized Library

The primary screening methods described above will yield a separate "pool"of candidate Zf proteins for each "Zf/subsite" pair. A key aspect of the CSPO strategy is that these "pools" can be recombined to produce a secondary library comprising variants that harbor fingers which have been partially optimized for binding to a desired subsite. For example, such a secondary library can comprise a range of multi-finger proteins composed of random combinations of the pools of fingers selected from the randomized fingers of the primary library. Thus, the secondary library can comprise multi-finger

22

proteins that, unlike the primary library, can potentially vary at all finger positions of the multi-finger proteins. Furthermore, the secondary library can comprise fingers with a range of binding affinities and specificities for their target subsite(s). The secondary library can then be used in a secondary screen, which is preferably conducted under conditions of high-stringency, to produce a multi-Zf polypeptide that binds with high affinity to the sequence of interest. Preferably, a new secondary library is synthesized for each new multi-finger protein to be produced.

The individual "pools" derived from the individual primary selections can be recombined using any one of a number of recombination techniques known in the art, such as described in, for example, Sambrook et al., Molecular Cloning; A Laboratory Manual 2d ed. (1989). Preferably, the individual "pools" derived from the individual primary selections are recombined using a PCR-mediated recombination method. More preferably still, the individual "pools" derived from the individual primary selections are recombined using the PCR-mediated recombination method outlined in Figure 2.

VII. Secondary Screening and Selection

For each gene of interest-specific multi-Zf protein to be produced, a single high-stringency secondary screen is preferred. In this screen, a partially optimized secondary library (such as described above) is screened against the sequence of interest, wherein the sequence of interest excludes "anchor" subsites. Thus, in the secondary screen, full-length assembled Zfs that bind to the sequence of interest can be identified. This is a key aspect of the present invention, as it means that there is no need to perform any post-selection assembly of individual Zfs or groups of Zfs. Such post-selection assembly is a common feature of other Zf selection methods. Post-selection assembly often introduces an uncontrollable element into the production of multi-finger proteins, as there is a possibility that the individually selected fingers will not function as predicted when assembled into the final multi-finger protein. Methods of the present invention advantageously allow for secondary selection of fully assembled Zfs, thereby accounting for potential finger position sensitivity.

In a preferred embodiment, the secondary selection is performed at high-stringency in order to isolate proteins that bind to their sequence of interest with high affinity. Mechanisms for controlling the stringency of selection reactions are known to

23

those of skill in the art and any such mechanism can be used.

VIII. Characterization of CSPO selected proteins

Recombinant Zf proteins identified using methods of the present invention can be further characterized after selection to ensure that they have the desired characteristics for their chosen use. Furthermore, the selected proteins can be tested using a different strategy than that used in the original selection, thereby controlling for the possibility of spurious or artifactual interactions specific to the selection system. For example, Zfs selected using a bacterial two-hybrid or phage-display system can be assayed for binding to their target sequence using an electrophoretic mobility shift assay or "EMSA" (Buratowski & Chodosh, in Current Protocols in Molecular Biology pp. 12.2.1-12.2.7). Equally, any other DNA binding assay known in the art could be used to verify the DNA binding properties of the selected protein.

Preferably, calculations of binding affinity and specificity are also made. This can be done by a variety of methods. The affinity with which the selected Zf protein binds to the sequence of interest can be measured and quantified in terms of its K_D. Any assay system can be used, as long is it gives an accurate measurement of the actual K_D of the Zf protein. In one embodiment, the K_D for the binding of a Zf protein to its target is measured using an EMSA

In a preferred embodiment, EMSA is used to determine the K_D for binding of the selected Zf protein both to the sequence of interest (i.e. the specific K_D) and to non-specific DNA (i.e. the non-specific K_D). Any suitable non-specific or "competitor" double stranded DNA known in the art can be used. Preferably, calf thymus DNA is used. The ratio of the specific K_D to the non-specific K_D can be calculated to give the specificity ratio. Zfs that bind with high specificity have a high specificity ratio. This measurement is very useful in deciding which of a group of selected Zfs should be used for a given purpose. For example, use of Zfs *in vivo* requires not only high affinity binding but also high-specificity binding. In a preferred embodiment, Zfs isolated using methods of the present invention have binding specificities higher than Zfs selected using other selection strategies (such a parallel selection, sequential selection and bipartite selection), and even more preferably, comparable or superior to those of naturally occurring multi-finger proteins, such as Zif268.

24

IX. Use of CSPO Selected Proteins.

The ultimate aim of producing a custom-designed Zf domain by CSPO is to obtain a Zf that can be used to perform a function. The Zf DBD can be used alone, for example to bind to a specific site on a gene and thus block binding of other DNA-binding domains. In a preferred embodiment, the Zf will be used in the construction of a recombinant transcription factor. Such a recombinant transcription factor can be used for a variety of purposes including regulation gene expression *in vivo* for the treatment of disease or regulation of gene expression either *in vivo* or *in vitro* for the purpose of studying gene function (i.e. functional genomics).

To generate a functional transcription factor from a CSPO-selected Zf, at least the Zf domain is fused to an effector domain. The effector domain can be associated with the Zf protein at any suitable position, including the C- or N-terminus of the Zf protein.

Common regulatory domains for addition to the Zf protein made using the methods of the invention include effector domains from transcription factors (activators, repressors, co-activators, co- repressors), silencers, nuclear hormone receptors, oncogene transcription factors (e.g., myc, jun, fos, myb, max, mad, rel, ets, bcl, myb, mos family members etc.); and chromatin associated proteins and their modifiers (e.g. methylases, demethylases, acetylases and deacetylases).

Kinases, phosphatases, and other proteins that modify polypeptides involved in gene regulation are also useful as regulatory domains for Zf proteins. Such modifiers are often involved in switching on or off transcription mediated by, for example, hormones. Kinases involved in transcription regulation are reviewed in Davis, Mol. Reprod. Dev. 42:459-67 (1995). Phosphatases are reviewed in, for example, Schonthal & Semin, Cancer Biol. 6:239-48 (1995).

Fusions of CSPO-selected Zfs to regulatory domains can be performed by standard recombinant DNA techniques well known to those skilled in the art, and as are described in, for example, basic laboratory texts such as Sambrook et al., Molecular Cloning; A Laboratory Manual 2d ed. (1989).

EXAMPLES

The following examples are provided to describe and illustrate, but not limit, the

25

claimed invention. Those of skill in the art will readily recognize a variety of non-critical parameters that could be changed or modified to yield essentially similar results. As described herein, proteins produced by methods of the present invention have greater affinity and specificity for their target sites than proteins produced by alternative strategies that do not account for both finger position sensitivity and combinatorial diversity.

Example 1

Construction of Multi-Finger Position-sensitivePrimary Library

Three different randomized "Primary Libraries" were constructed, each library comprising three fingers, one of which was variable/randomized and two of which were "anchored." In "Primary Library 1" the N-terminal Zf (Zf 1) was randomized while Zf 2 and Zf 3 were held constant. In "Primary Library 2" the middle Zf (Zf 2) was randomized while Zf 1 and Zf 3 were "anchored." In "Primary Library 3" the C-terminal Zf (Zf 3) was randomized while Zf 1 and Zf 2 were "anchored.". These three libraries were constructed essentially as previously described by Joung et al. (Joung et al., (2000) Proceedings of the National Academy of Sciences (USA) 97: 7382), with two exceptions. The first exception was that different finger positions were randomized for each library made (i.e. Primary Library 1, Primary Library 2, and Primary Library 3). The second exception was that the 24 codons used to randomize amino acid residues in the recognition helix, encoded only 16 of the possible 20 amino acids. The excluded amino acids were phenylalanine, tyrosine, tryptophan and cysteine.

Example 2

Construction of Position-sensitive Target sites
for Selection of Zf Polypeptides that Bind to the BCR-ABL Gene

Target sites were synthesized as oligonucleotides and introduced just upstream of the weak test promoter in the bacterial two-hybrid system, as described in Joung et al., (2000) Proceedings of the National Academy of Sciences (USA) 97:7382.

Example 3

Construction of a Partially Optimized Secondary Library

The CSPO protocol (illustrated in Figure 1) was designed so that "pools" of Zfs that bind with low affinity to their respective subsites in the primary selection could be

isolated and recombined to generate a "Secondary Library." Such secondary libraries were produced using PCR-mediated recombination, according to the method illustrated in Figure 2.

Example 4

Quantification of Target Binding Affinity and Specificity

Zf proteins selected using CSPO were characterized to determine the affinity and specificity with which they bound to their target sites. DNAs encoding selected Zfs were isolated. In order to produce the encoded Zf protein *in vitro*, a commercially available *in vitro* transcription/translation system (Expressway^{TM}, Invitrogen) was used. The binding of the *in vitro* transcribed/translated Zf proteins to their target sites was measured assayed using electrophoretic mobility shift assays (EMSAs). EMSAs were performed with a constant concentration of the DNA target sites and a range of concentrations of the Zf protein being tested. Thus, by quantifying the amount of the Zf protein bound to the target at each Zf protein concentration, it was possible to obtain a measure of the K_D for binding of the Zf protein to its target. Figure 3 shows the data EMSA and K_D data obtained for a Zf selected for binding to an HIV-1 promoter sequence using the CSPO strategy.

Figure 4 shows the results obtained when a similar EMSA was performed in which the Zf protein concentration was held constant and the concentration of non-specific competitor DNA (calf thymus DNA) was varied. By quantifying the amount of the Zf protein bound to the target at each non-specific DNA concentration, it was possible to obtain a measure of the K_D for binding of the Zf protein to non-specific DNA. Figure 4 shows the EMSA and non-specific K_D data obtained for a Zf selected for binding to an HIV-1 promoter sequence using the CSPO strategy.

Example 5

Selection of Zf Polypeptides with High Affinity and Specificty for the BCR-ABL Gene

Choo et al. (1994, Nature 372:642) have previously described the use of the parallel selection strategy to select a recombinant three-finger Zf protein that binds specifically to a unique 9 bp region of a BCR-ABL fusion oncogene. This recombinant 3-finger protein has the amino acid sequence DRSSTR QGGNVR QAATQR (SEQ ID NO:8) in the recognition helices of finger 1, 2, and 3, respectively, and binds to the BCR-

27

ABL target sequence GCA GAA GCC (SEQ ID NO:9) (Figure 5).

In the present example, CSPO was used in conjunction with a bacterial two-hybrid screening system, to select recombinant Zfs that bind to the same 9 bp BCR-ABL target sequence, i.e. GCA GAA GCC (SEQ ID NO:9).

Twelve recombinant Zf proteins, termed BCAB1 through BCAB12, were selected (Figure 6). Each of these Zf proteins differed in sequence from the Zf protein isolated by Choo et al. (referred to as "wild-type" for the purposes of this example only). The two strongest binders, BCAB1 and BCAB7, were further characterized and compared to the wild-type protein. Dissociation constants (K_D) for binding to the BCR-ABL target sequence were measured and quantified using electrophoretic mobility shift assays (EMSAs). Specificity of binding was determined by comparing the K_D for binding to the BCR-ABL target sequence to the K_D for binding to non-specific competitor DNA. Figure 7 shows the K_D s for specific and non-specific binding and the calculated "specificity ratios." The results of this analysis demonstrate that both BCAB1 and BCAB7 bind with high affinity to the BCR-ABL target sequence, and furthermore, that they bind with higher specificity than the "wild-type" protein.

Thus, using the context-sensitive parallel optimization strategy of the present invention, recombinant Zfs with desirable target binding characteristics for this BCR-ABL target sequence, have been identified.

<div align="center">Example 6</div>

<div align="center">Selection with the erb-B2 Target Site</div>

Beerli et al. (1998, Proceedings of the National Academy of Sciences (USA) 95:14628) have previously described use of a parallel selection strategy to select a recombinant three-finger Zf protein that binds specifically to a 9 bp site in the human erb-B2 gene. This recombinant 3-finger protein has the amino acid sequence RKDSVR QSGDRR DCRDAR (SEQ ID NO:10) and binds to the erb-B2 sequence GCC GCA GTG (SEQ ID NO:11) (Figure 5). In the present example, CSPO was used in conjunction with a bacterial two-hybrid screening system to select recombinant Zfs that bind to the same 9 bp erb-B2 target site, i.e. GCC GCA GTG (SEQ ID NO:11).

Twelve recombinant Zf proteins, termed EB1 through EB12, were selected (Figure 8). Each of these Zf proteins differed in sequence from the Zf protein isolated by

<div align="center">28</div>

Beerli et al. (referred to as "wild-type" for the purposes of this example only). The two strongest binders, EB3 and EB11, were further characterized and compared to the "wild-type" protein. Dissociation constants (K_D) for binding to the erb-B2 target sequence were measured and quantified using EMSAs. Specificity of binding was determined by comparing the K_D for binding to the erb-B2 target sequence to the K_D for binding to non-specific competitor DNA. Figure 9 shows the K_D s for specific and non-specific binding and the calculated "specificity ratios." The results of this analysis demonstrate that both EB3 and EB11 bind to the erb-B2 target with higher affinity than the "wild-type" protein, and furthermore, that they bind with specificity similar to that of the "wild-type" protein. The specificity ratios for the selected proteins are greater than that of the "wild-type" protein.

Thus, using the context-sensitive parallel optimization strategy of the present invention, recombinant Zfs with desirable target binding characteristics for this erb-B2 target sequence, have been identified.

Example 7

Selection with the HIV Promoter

Isalan et al. (2001, Nature Biotechnology19: 656) have previously described the use of the bipartite selection strategy to select a recombinant three-finger Zf protein that binds specifically to a 9 bp site in the human immunodeficiency virus 1 (HIV-1) promoter. This recombinant 3-finger protein has the amino acid sequence ASADTR NRSDSR TSSNKK (SEQ ID NO:12) and binds to the HIV-1 promoter target sequence GAT GCT GCA (SEQ ID NO:13) (Figure 5).

In the present example, CSPO was used in conjunction with a bacterial two-hybrid screening system, to select recombinant Zfs that bind to the same 9 bp. HIV-1 promoter target sequence GAT GCT GCA (SEQ ID NO:13).

Twelve recombinant Zf proteins, termed HP1 through HP12, were selected (Figure 10). Each of these Zf proteins differed in sequence from the Zf protein isolated by Isalan et al. (referred to as "wild-type" for the purposes of this example only). The two strongest binders, HP6 and HP12, were further characterized. Dissociation constants (K_D) for binding to the HIV-1 promoter sequence were measured and quantified using EMSAs. Specificity of binding was determined by comparing the K_D for binding to the

HIV-1 promoter sequence to the K_D for binding to non-specific competitor DNA. Figure 11 shows the K_Ds for specific and non-specific binding and the calculated "specificity ratios." The results of this analysis demonstrate that both HP6 and HP12 bind to the HIV-1 promoter with high affinity and specificity. It was not possible to compare the target binding affinities and specificities of HP6 and HP12 to those of the "wild-type" protein in the present study.

Thus, using the CSPO strategy of the present invention, recombinant Zfs with desirable target binding characteristics for the HIV-1 promoter have been identified.

While a preferred form of the invention has been shown in the drawing and described in some detail, variations in the preferred form will be apparent to those skilled in the art and thus the invention should not be construed as limited to the specific form shown and described, but instead is as set forth in the following claims.

CLAIMS

We claim:

1. A method of selecting a multi-finger zinc finger polypeptide that recognizes a sequence of interest comprising a target site having at least one subsite, said method comprising the steps of:

    a) first obtaining primary libraries comprising polypeptides having one variable finger and at least one anchor finger, wherein said variable finger corresponds to a zinc finger of said multi-finger zinc finger polypeptide;

    b) incubating said primary libraries with said target site under conditions sufficient to form binding complexes;

    d) isolating pools comprising nucleic acid sequences encoding polypeptides, wherein said polypeptides comprise said binding complexes;

    e) recombining said pools to produce a secondary library;

    f) incubating said secondary library with the sequence of interest under conditions sufficient to form a high-affinity binding complex; and

    g) isolating nucleic acid sequences encoding multi-finger zinc finger polypeptides, wherein said polypeptides comprise said high-affinity binding complexes.

2. The method of claim 1, wherein the multi-finger zinc finger polypeptide comprises at least two zinc fingers.

3. The method of claim 2, wherein the multi-finger zinc finger polypeptide comprises three zinc fingers.

4. The method of claim 1, wherein a subsite of the target site comprises 3 bp.

5. The method of claim 1, wherein the target site comprises at least two subsites.

6. The method of claim 5, wherein the target site comprises three subsites.

31

7.   The method of claim 1, wherein the primary libraries comprise polypeptides having at least one anchor finger that corresponds to a zinc finger polypeptide.

8.   The method of claim 1, wherein the anchor finger(s) bind to target subsites with low affinity.

9.   The method of claim 7, wherein the zinc finger polypeptide is selected from the group consisting of Zif268, tramtrack, GLI and TFIIA.

10.   The method of claim 8, wherein the zinc finger polypeptide is Zif268.

11.   The method of claim 9, wherein the zinc finger polypeptide is a phage-selected derivative of Zif268.

12.   The method of claim 11, wherein the phage-selected derivative of Zif268 comprises sequences selected from the group consisting of SEQ ID NO:2 (DRSSLTR, finger 1), SEQ ID NO:3 (QGGNLVR, finger 2) and SEQ ID NO:4 (QAATLQR, finger 3) and combinations thereof.

13.   The method of claim 1, wherein the primary library comprises polypeptides having two or more anchor fingers.

14.   The method of claim 1, wherein the variable finger is derived from a zinc finger polypeptide.

15.   The method of claim 14, wherein the zinc finger polypeptide is selected from the group consisting of Zif268, tramtrack, GLI and TFIIA.

16.   The method of claim 15, wherein the zinc finger polypeptide is Zif268.

17.   The method of claim 14, wherein the zinc finger polypeptide is a phage-selected derivative of Zif268.

18.   The method of claim 17, wherein the phage-selected derivative of Zif268 comprises sequences selected from the group consisting of SEQ ID NO:2 (DRSSLTR, finger 1), SEQ ID NO:3 (QGGNLVR, finger 2) and SEQ ID NO:4 (QAATLQR, finger 3) and combinations thereof.

19.   The method of claim 1, wherein the variable finger comprises six randomized amino acid residue positions within an alpha helix.

20.   The method of claim 19, wherein the randomized amino acid residue positions within the alpha helix are -1, +1,+2, +3, +5 and +6.

21.   The method of claim 19, wherein between 16 to 20 amino acids are represented at each randomized position.

22.   The method of claim 21, wherein between 16 to 19 amino acids are represented at each randomized residue position.

23.   The method of claim 22, wherein 16 amino acids are represented at each randomized residue position.

24.   The method of claim 1, wherein the target site comprises the same number of base pairs as the sequence of interest.

25.   The method of claim 24, wherein the target site comprises two or more subsites.

26.   The method of claim 25, wherein one subsite has a sequence identical to the target site and the remaining subsite(s) have sequences that bind to the anchor finger(s).

27. The method of claim 26, wherein the remaining subsite(s) sequences selected from the group consisting of SEQ ID NO:5 (GCC subsite 1), SEQ ID NO:6 (GAA subsite 2) and SEQ ID NO:7 (GCA subsite 3) and combinations thereof.

28. The method of claim 1, wherein the primary libraries are expressed *in vitro*.

29. The method of claim 1, wherein the primary libraries are expressed in expression systems selected from the group consisting of eukaryotic, prokaryotic and viral expression system.

30. The method of claim 29, wherein the primary libraries are expressed in bacteria.

31. The method of claim 1, wherein incubation of the primary libraries is performed *in vitro*.

32. The method of claim 1, wherein incubation of the primary libraries is performed within a prokaryotic or eukaryotic cell.

33. The method of claim 32, wherein the incubation is performed within a bacterial cell.

34. The method of claim 1, wherein the isolated pools of nucleic acid sequences are recombined to produce a secondary library by PCR-mediated recombination.

35. The method of claim 1, wherein the secondary library is expressed *in vitro*.

36. The method of claim 1, wherein the secondary library is expressed in an expression system selected from the group consisting of a eukaryotic, prokaryotic and viral expression system.

37. The method of claim 36, wherein the secondary library is expressed in bacteria.
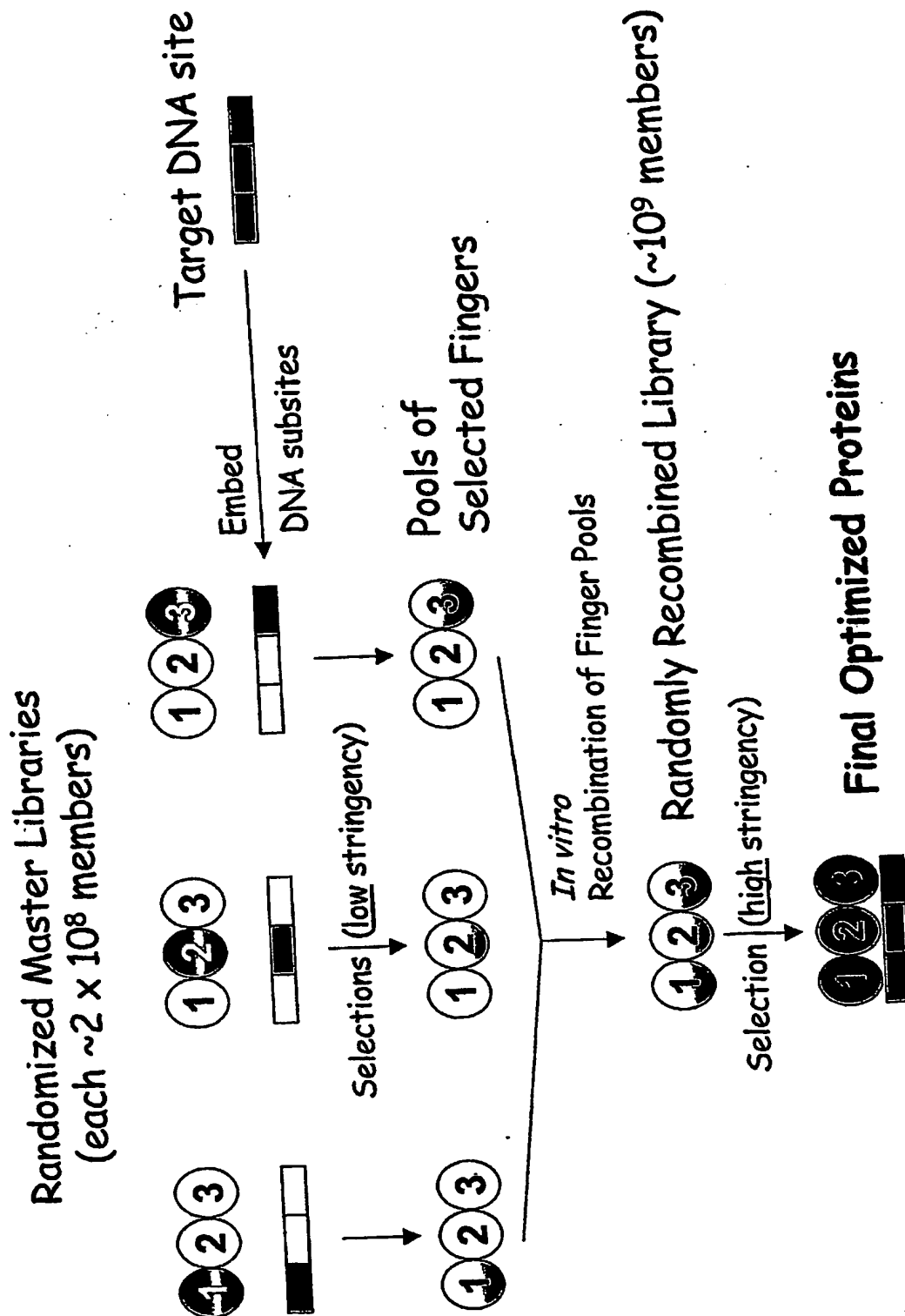
38.   The method of claim 1, wherein incubation of the secondary library is performed *in vitro.*

39.   The method of claim 1, wherein incubation of the secondary library is performed within a prokaryotic or eukaryotic cell.

40.   The method of claim 39, wherein the incubation of the secondary library is performed within a bacterial cell.

41.   A kit for selecting a multi-finger zinc finger polypeptide according to any one of the preceding claims.

42.   A multi-finger zinc finger polypeptide selected according to any of the preceding claims.

43.   A method of regulating gene expression comprising contacting a multi-finger zinc finger polypeptide according to claim 42 with a gene of interest in an expression system.

44.   A multi-finger zinc finger polypeptide according to claim 42, wherein the multi-finger zinc finger polypeptide is fused to a regulatory or effector domain to generate a recombinant transcription factor.

45.   A method of regulating gene expression comprising contacting a multi-finger zinc finger polypeptide according to claim 44 with a gene of interest in an expression system.

46.   A method of selecting a multi-finger zinc finger polypeptide that recognizes a sequence of interest comprising a target site having at least one subsite, said method comprising the steps of:

a) incubating primary libraries comprising polypeptides having one variable finger and at least one anchor finger, wherein said variable finger corresponds to a zinc finger of said multi-finger zinc finger polypeptide, with said target site under conditions sufficient to form binding complexes;

b) isolating pools comprising nucleic acid sequences encoding polypeptides, wherein said polypeptides comprise said binding complexes;

c) recombining said pools to produce a secondary library;

d) incubating said secondary library with the sequence of interest under conditions sufficient to form a high-affinity binding complex; and

e) isolating nucleic acid sequences encoding multi-finger zinc finger polypeptides, wherein said polypeptides comprise said high-affinity binding complexes.

ABSTRACT

The present invention relates to methods of identifying multi-finger Zf polypeptides that bind with high affinity and specifity to multi-subsite target sequences. The invention provides an efficient selection strategy that allows pre-assembled multi-finger polypeptides to be selected for binding to a desired sequence of interest while also retaining full combinatorial diversity in the Zf libraries used. Zf polypeptides identified using the methods described herein have affinity and specificity for their target sites that is superior to those produced by alternative methods.

Figure 1

Context-Sensitive Parallel Optimization

Randomized Master Libraries
(each ~2 x 10⁸ members)

Target DNA site

Embed
DNA subsites

Selections (low stringency)

Pools of
Selected Fingers

In vitro
Recombination of Finger Pools

Randomly Recombined Library (~10⁹ members)

Selection (high stringency)

Final Optimized Proteins

Figure 2

Construction of Randomly Recombined Libraries

DNA sequences of finger pools selected
in the primary selections

PCR Amplification of Selected Finger Pools with finger-specific primers

Random recombination of
fingers pools by
overlap-mediated PCR

PCR amplification of full-length
recombined product using end primers

Randomly recombined
library of finger pools

Finger 1   Finger 2   Finger 3

Finger 1   Finger 2   Finger 3

Figure 3

Quantifying Affinity of ZFPs

- Purification of ZFPs -- used coupled in vitro transcription/ translation system (Expressway from Invitrogen)
  - faster and avoids toxicity issues associated with overexpression in *E. coli*
  - isolated as a MBP fusion using affinity purification
  - 1 day for expression + purification

- Affinity of a given ZFP for its target DNA site can be quantified by the dissociation constant (K_d^{spec})

K_d Experiment (HP6)

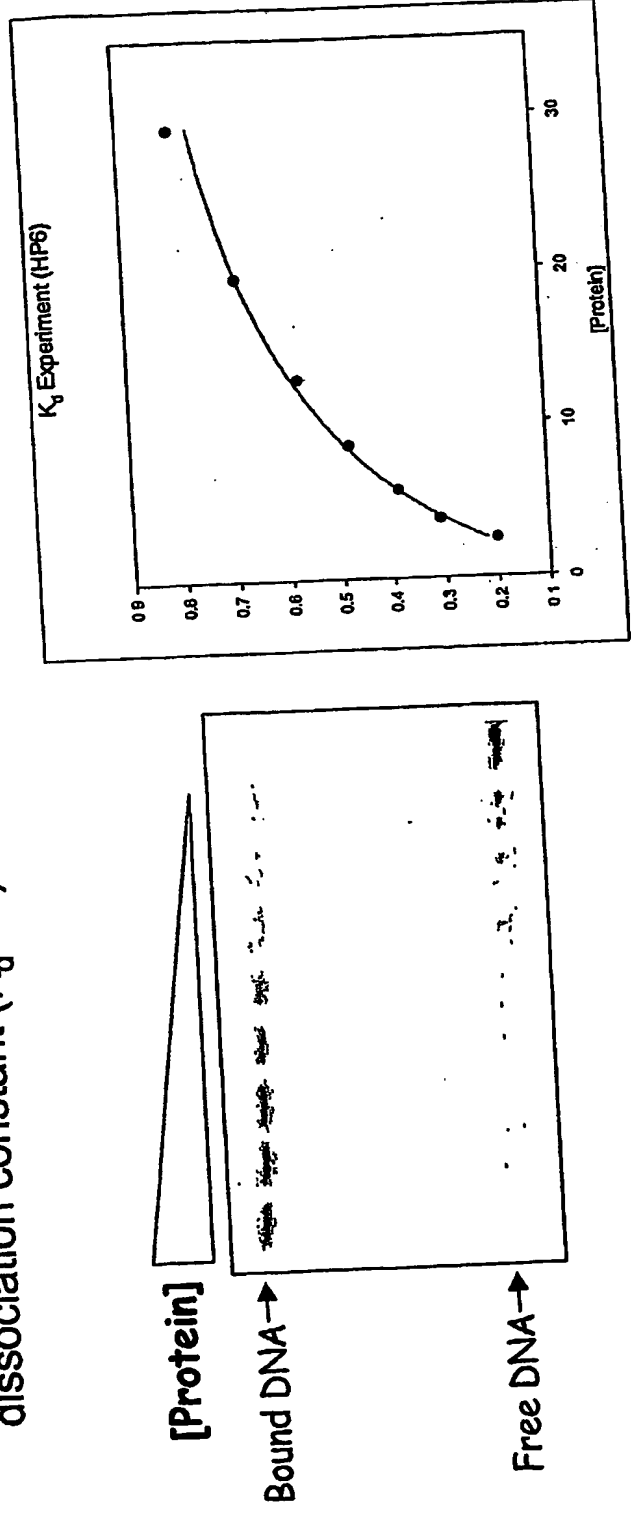[Protein]

[Protein]

Bound DNA→

Free DNA→

Figure 4

Characterizing Specificity of ZFPs

- Another approach: calculate Kd of protein for non-specific DNA (K_d^{NS})

- Compare this value to the Kdspec to assess specificity of the protein

Figure5

Validating Context-Sensitive Parallel Optimization

DNA target sites:

Previously selected Multi-finger Proteins:

① ② ③

BCR-ABL	5'GCAGAAGCC3'	DRSSTR	QGGNVR	QAATQR*
erb-B2	5'GCCGCAGTG3'	RKDSVR	QSGDRR	DCRDAR*
HIV promoter	5'GATGCTGCA3'	ASADTR	NRSDSR	TSSNKK#

*identified by Modular Parallel Optimization
#identified by Bipartite Optimization

Figure 6

Selections for the BCR-ABL site

CCG   AAG   ACG ← DNA target site
F1   F2   F3

	-112356	-112356	-112356
"wild-type"	DRSSTR	QGGNVR	QAATQR
BCAB1	DSPTRR	QGANRR	QANTQR
BCAB2	DSPTRR	QGANRR	QANTQR
BCAB3	DSPTRR	QGANRR	QANTQR
BCAB4	DESTRR	QSPNRR	QRNTQR
BCAB5	DSPTRR	QGGNRR	QGNTTR
BCAB6	DSPTRR	QNPNVR	QRNTQR
BCAB7	DSPTRR	QRTMIR	QRNTQR
BCAB8	DRPTRR	QGGNVR	QANTQR
BCAB9	DSPTRR	QNTHNR	QANTQR
BCAB10	DSPTRR	QKPMDR	QGNSIR
BCAB11	DSPTRR	QSTNNR	QGNSNR
BCAB12	DSPTRR	QNVATR	QRNTQR

fold-activation

Figure 7

In vitro characterization of BCR-ABL ZFPs

Protein	Sequence	K_d^{spec} (pM)	$K_d^{non-spec}$ (nM)	Specificity ratio	# of DNA bases specified
"wt"	DRSSTR QGGNVR QAATQR	28(±3.9)	55(±12)	1,980	~5.5
BCAB 1	DSPTRR QGANRR QANTQR	78(±13)	2100(±270)	27,000	~7.4
BCAB 7	DSPTRR QRTNIR QRNTQR	60(±8.5)	1300(±97)	23,000	~7.2
Zif268		8.1(±1.8)	1000(±120)	130,000	~8.5

Figure 8

Selections for the erb-B2 site

GTG ACG CCG → DNA target site
  F1   F2   F3

-1 1 2 3 5 6 -1 1 2 3 5 6 -1 1 2 3 5 6

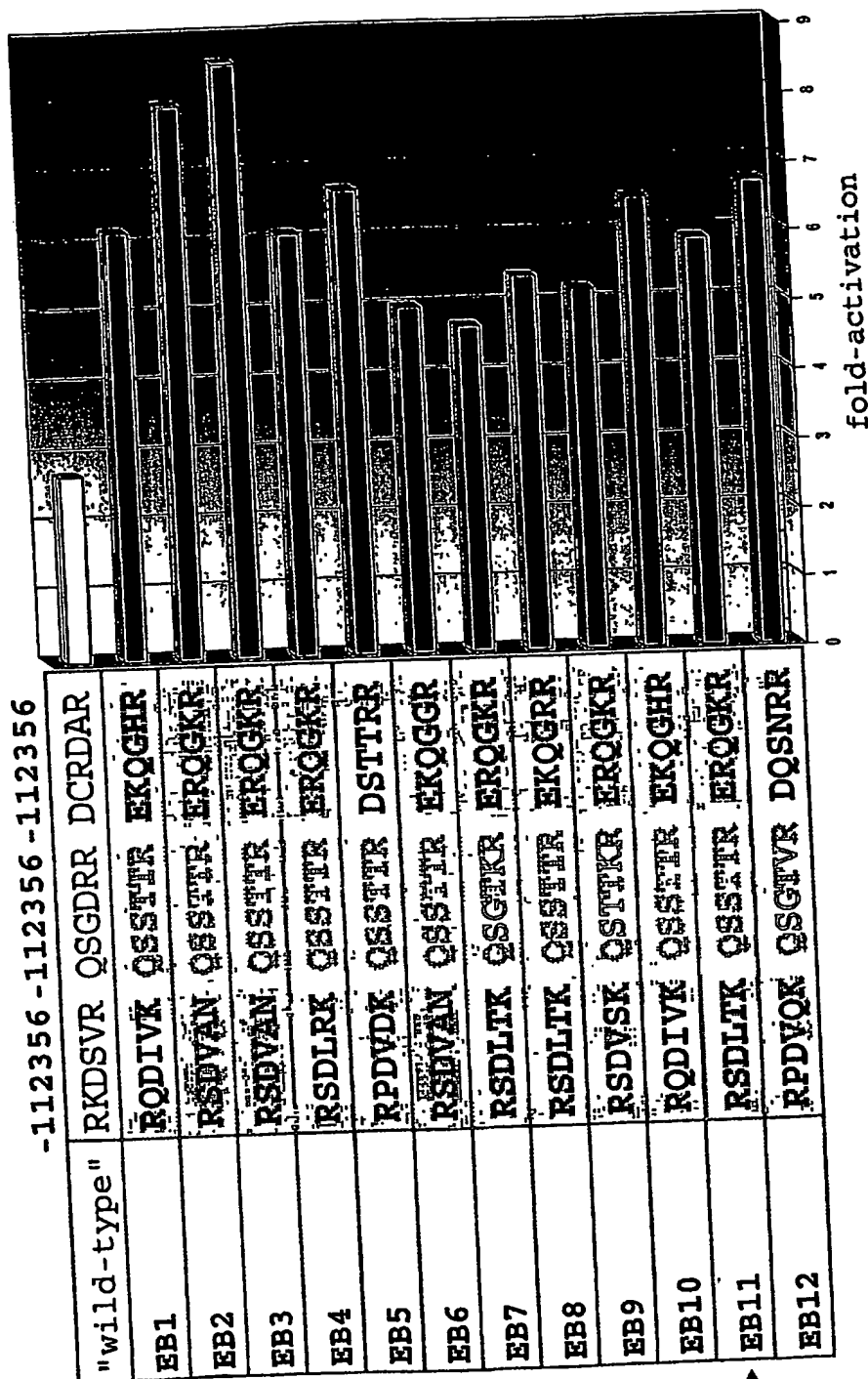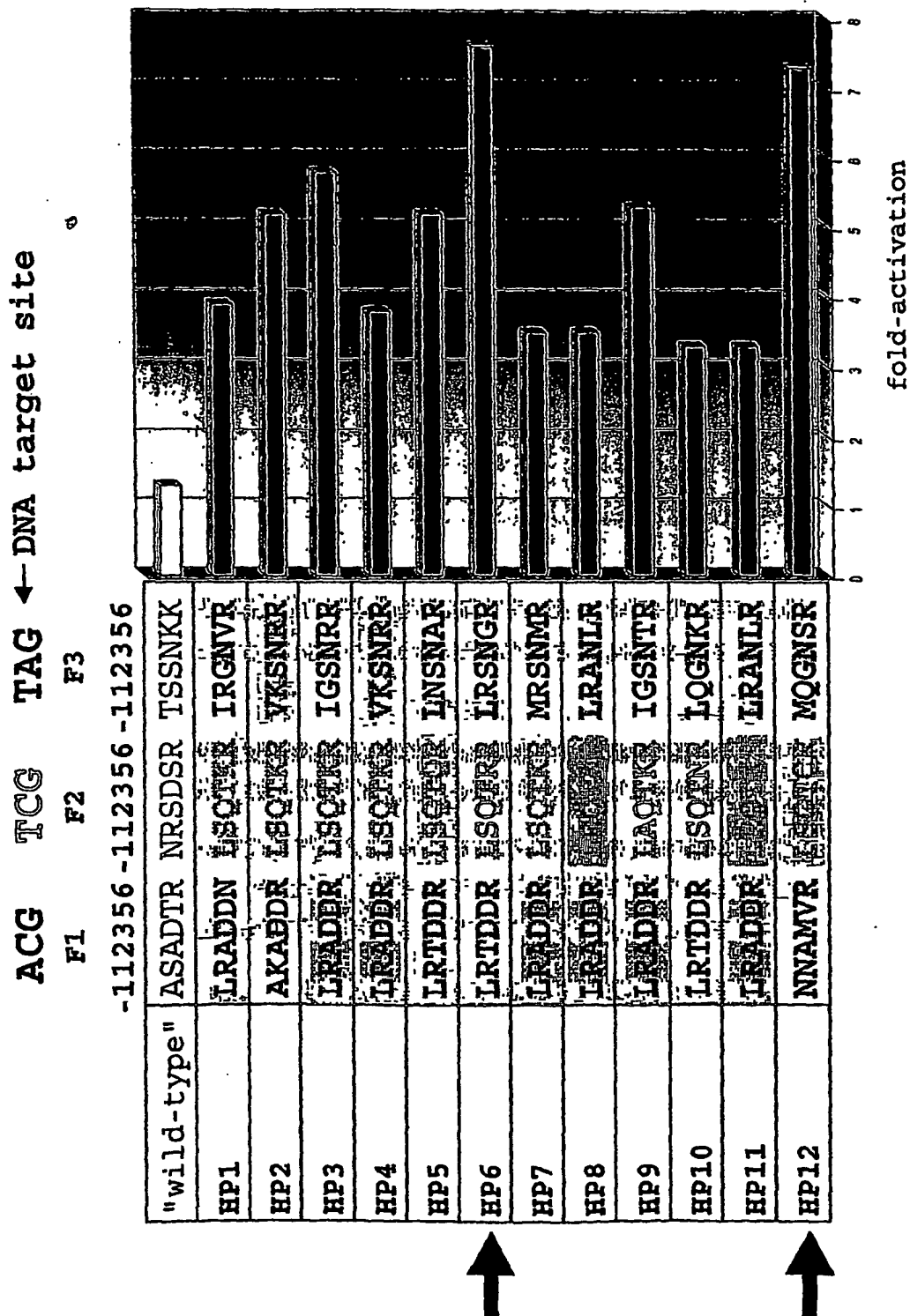|	F1	F2	F3
"wild-type"	RKDSVR	QSGDRR	DCRDAR
EB1	RQDIVK	QSSTTR	EKQGHR
EB2	RSDVAN	QSSTTR	ERQGKR
EB3	RSDVAN	QSSTTR	ERQGKR
EB4	RSDLRK	QSSTTR	ERQGKR
EB5	RPDVDK	QSSTTR	DSTTRR
EB6	RSDVAN	QSSTTR	EKQGGR
EB7	RSDLTK	QSGEKR	ERQGKR
EB8	RSDLTK	QSSTTR	EKQGRR
EB9	RSDVSK	QSTTKR	ERQGKR
EB10	RQDIVK	QSSTTR	EKQGHR
EB11	RSDLTK	QSSTTR	ERQGKR
EB12	RPDVQK	QSGTVR	DQSNRR

fold-activation

Figure 9

In vitro characterization of erb-B2 ZFPs

Protein	Sequence	K_d^{spec} (pM)	$K_d^{non\text{-}spec}$ (nM)	Specificity ratio	# of DNA bases specified
"wt"	RKDSVR QSGDRR DCRDAR	150 (±23)	1000 (±120)	6,700	~6.4
EB 3	RSDVAN QSSTTR ERQGKR	31 (±3.1)	1100 (±15)	35,000	~7.5
EB 11	RSDLTK QSSTTR ERQGKR	65 (±3.9)	1100 (±81)	17,000	~7.0
zif268		8.1 (±1.8)	1000 (±120)	130,000	~8.5

Figure 10

Selections for the HIV promoter site

ACG	TCG	TAG	← DNA target site
F1	F2	F3	
-112356	-112356	-112356	

	F1	F2	F3
"wild-type"	ASADTR	NRSDSR	TSSNKK
HP1	LRADDN	LSQTLR	IRGNVR
HP2	AKABDR	LBQTKR	VKSNRR
HP3	LRADDR	LSQTHR	IGSNRR
HP4	LRADDR	LSQPHR	VKSNRR
HP5	LRTDDR	LSGHDM	LNSNAR
HP6	LRTDDR	LSQHRR	LRSNGR
HP7	LRADDR	LSQTNR	MRSNMR
HP8	LRADDR		LRANLR
HP9	LRADDR	LAQHKR	IGSNTR
HP10	LRTDDR	LSQTNR	LQGNKR
HP11	LRADDR		LRANLR
HP12	NNAMVR		MQGNSR

fold-activation

Figure 11

In vitro characterization of HIV Promoter ZFPs

Protein	Sequence	K_d^{spec} (pM)	$K_d^{non\text{-}spec}$ (nM)	Specificity ratio	# of DNA bases specified
"wt"	ASADTR NRSDSR TSSNKK	Unable to calculate (does not bind in vitro)			
HP6	LRTDDR LSQTRR LRSNGR	9.3 (±1.2)	820 (±74)	87,000	~8.2
HP12	NNAMVR LSQTQR MQGNSR	9.3 (±0.39)	180 (±8.8)	19,000	~7.1
zif268		8.1 (±1.8)	1000 (±120)	130,000	~8.5